DOCUMENT RESUME

ED 392 820 TM 024 468

AUTHOR Webster, William J.
TITLE What Principals Need To Know about Measurement in the Era of Site-Based Management.
PUB DATE Apr 95
NOTE 14p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; Budgeting; Curriculum Development; Decision Making; *Educational Assessment; Elementary Secondary Education; Evaluation Methods; Measurement Techniques; *Performance Based Assessment; Personnel Selection; Portfolios (Background Materials); *Principals; Reliability; *School Based Management; Testing Programs; Test Use; Validity

ABSTRACT

The current trend toward site-based management in education, with its accompanying shift toward decision making at the individual school level, means new responsibilities and roles for principals. Principals are now required to be intimately involved in budgeting, personnel recruitment and hiring, selection of curricula and textbooks, and many other tasks. To meet accountability requirements, principals must become competent in measurement and the design of practical evaluations. The competencies principals must demonstrate in assessment and accountability under site-based management can be divided into: (1) basic measurement concepts, such as reliability, validity, test types, and performance assessment, including portfolios; (2) knowledge about the use of test data to improve instruction, including student test-taking skills and the purposes of testing; (3) basic evaluation concepts; and (4) the characteristics of a good testing program and criteria for judging assessment quality. (Contains 13 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 392 820

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

What Principals Need to Know About Measurement In The Era of Site-Based Management[1]

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
WILLIAM J. WEBSTER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

William J. Webster
Dallas Public Schools

The current trend in the administration of the public schools is toward site-based management. This means different things to different administrators and Boards of Education, but, for the purposes of this paper, it is defined as the shift of most authority and responsibility for educational decision-making to the local campus level. Although this shift is often accompanied by a designed increase in parental, community, and teacher involvement, the accountability is usually focused primarily on the principal. Thus while many interest groups often share in the decision-making process, it is the principal that must orchestrate the process so that the end result is positive for the school's primary clients, the students.

Site-based management requires the manager to fill a role that is unfamiliar and that the principal is often unprepared to fill. Whereas the principal's role was formally one of managing resources that were generally provided with little input, the new role requires that principals be intimately involved in budget preparation, personnel recruitment and hiring, instructional program selection, textbook selection, and a myriad of other tasks related to delivering an effective instructional program. With this role change also comes increased accountability.

Most principals must contend with state and local accountability systems. These systems are often seriously flawed resulting in administrators being held to false or inappropriate standards. Many times these standards are used to make career decisions about principals. Even in situations where this is not the case, managers must be able to use test information and other data to make intelligent decisions about instructional program design and modification. Therefore, in order to be able to competently function in their new role, principals must be minimally competent in specific areas of measurement and practical evaluation design. That is,

---

1 Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, California, April 17-21, 1995

TM024468

BEST COPY AVAILABLE

measurement concepts are necessary but not sufficient competencies in the area of assessment and the related area of evaluation.

A brief review of the literature suggests that several professional organizations have identified criteria for educational administrators in the area of student assessment. These include the American Association of School Administrators (1993) and the National Association of Elementary School Principals (1991). While these sources address very broad competencies, they do not provide specifics in terms of the knowledge and skills needed by principals to properly use assessment in the management of schools. Popham and Hambleton (1990) have specifically addressed measurment concepts and competencies required by principals while Impara, et. al. (1993a, 1993b, 1994) have done a great deal of work in the area of administrators' and teachers' knowledge and attitudes about assessment. This paper addresses specific assessment and evaluation competencies required to successfully operate a school that is site-based managed. It is written from the perspective of a person who works with principals daily on a myriad of assessment and accountability issues. These competencies are divided into four general areas: basic measurement concepts, knowledge about the use of test data in improving instruction, basic evaluation concepts, and the characteristics of a good testing program.

Basic Measurement Concepts

Based on twenty-five years of interaction with principals on measurement issues, the amount that the average principal doesn't know about testing and measurement is alarming. This is largely because their graduate preparation programs never required them to learn much about assessment and even less about evaluation. This lack of knowledge is particularly relevant because unreliable and invalid tests are often severely impacting their lives. While not suggesting that principals become measurement experts, there are some minimum competencies that principals must possess if they are to have any influence in determining their own destinies. They at least need to know what questions to ask when confronted with test data. This is particularly true given the methodological sophistry that is an integral part of many state testing programs. Important measurement concepts include:

Reliability. It is important for principals to be familiar with the concept of consistency in measurement and the related concept of measurement error. Questions that ought to be asked by practitioners on a regular basis are "what is the standard error of measurement associated with that test score?" and "how reliable is this test for my student population?" Basic facts that

principals should know about reliability include the fact that the level of reliability depends in part on the length of the test and the variance of scores in the group studied, that a reliability coefficient specifies what proportion of test variance is non-error variance, that tests may measure more or less reliably at different points in the distribution, and that in order to be valid a test must first be reliable. Principals also need to be conceptually familiar with the various methods of estimating reliability (alternate form, stability, internal consistency).

Validity. While most principals are familiar with the concept of content validity, the importance of predictive and concurrent validity to the specific use of many tests is largely unknown or ignored. Many states have graduation criteria that include passing tests without any concern for the predictive validity of those tests. Principals should also be aware that if a particular test produces results that are unique when it purports to measure concepts that are common and measured by other tests or instruments, there may be a problem with concurrent validity. Construct validity is probably a bit esoteric, but principals should be conceptually familiar with the common methods of estimating predictive, concurrent, and construct validity.

Criterion-referenced Testing. Principals should understand the concept of criterion-referenced testing to the point where they understand that the primary purpose of criterion-referenced tests is to provide information about student performance relative to specific standards and that criterion-referenced testing should be used as an integral part of instruction. To this end they must be able to identify standard-setting issues and be knowledgeable of the common empirical and non-empirical methods used to establish standards. Criterion-referenced tests should be designed to provide sufficient information about students to allow for the diagnosis of individual student strengths and weaknesses. Most teacher made tests are criterion-referenced tests and principals must know the basic principles of criterion-referenced test design and construction in order to help teachers to optimally use their tests.

Norm-referenced Testing. One of the forms of testing that is most misunderstood by principals is norm-referenced testing. Since norm-referenced tests are often relied on heavily in many testing programs, it is essential that principals understand their uses and limitations. The principal must be familiar with the derivation and interpretation of a number of derived scores such as percentiles, stanines, grade equivalents, and normal curve equivalents. They must understand the importance of the norm group to the interpretation of test data. They must also have a basic understanding of the normal curve and how the normal curve and the norm group effect the derived score. Finally, they must be able to use the aforementioned information about tests to interpret individual and group test scores.

Scaling. One of the most misunderstood areas of testing by practitioners is that of scaling. Not only do they often have difficulty distinguishing between a percent and a percentile, many ascribe almost metaphysical properties to the 70% criterion. It is almost as if the statistic 70% has a life of its own, independent of the difficulty level of the test or of the persons being tested. Principals must have a rudimentary knowledge of scaling and item difficulty, and of the relationship between test and item difficulty and percentage of items correct.

Performance Testing. Principals should have a clear understanding of the issues involved with performance testing and of the impact that performance testing can have on the curriculum and subsequently on other test scores. It should be understood that while there is a problem with the reliability of performance tests, those types of exercises should be an integral part of the instructional delivery system and of teacher-made tests.

Portfolio Analysis. Principals should be familiar with the basic techniques of portfolio analyses and their use in student and teacher evaluation.

The Importance of Test Administration Procedures. It is essential that principals understand the importance of appropriate test administration procedures to the interpretation of test scores. Test scores and their interpretation assume certain standard administrative procedures. Without these procedures test scores are of extremely limited utility. Student scores are, in part, a function of test administration procedures.

In the area of basic measurement concepts, there are ten general areas with which principals must be familiar. These include a rudimentary knowledge of reliability, validity, criterion-referenced testing, norm-referenced testing, scaling, performance testing, portfolio analyses, and the importance of test administration procedures. While not becoming measurement experts, it is essential that principals become knowledgeable consumers of testing information. If there were just one publication that I would have principals read in this area, it would be the *Standards for Educational and Psychological Testing (1985)*. They would then be familiar with some of the major issues regarding testing and test use.

The Use Of Test Data In Improving Instruction

Once the educational leader is relatively sure that available data are reliable and valid, appropriately scaled, and obtained through valid test administration procedures, the ability to use these data to improve instruction becomes crucial. The principal must be able to link curriculum content and instructional strategies to assessment information. The principal must also be able to guide teachers in accomplishing this in their classrooms. There are a number of skills involved in this activity. They include the ability to embed assessment into classroom instruction so that teachers have timely and accurate feedback on student performance; the ability to understand and interpret skills analyses at the individual student and classroom level; the ability to interpret norm-referenced and criterion-referenced test scores; the ability to make sense of conflicting assessment information; and, most important, the aforementioned ability to link instructional strategies and curriculum modifications to assessment information. The last skill requires a knowledge of curriculum and instructional strategies and is the point at which many educational leaders fail.

Test-taking Skills. While teaching test-taking skills is a worthwhile activity for students who don't know how to take standardized tests, it is crucial that educators also concentrate on improving the curriculum. There is a significant difference between teaching test-taking skills and teaching the test. Many principals are under the impression that, by drilling and practicing similar test items, test scores can be significantly raised. The Dallas Public Schools has overwhelming evidence that, while this strategy may work at the lower end of the distribution (particularly when scores are around chance), it fails miserably once you have moved the majority of your students within a standard deviation of the mean of the test (Mendro, et. al., 1994).

Purposes of Testing. Finally, it is important that the educational leader be cognizant of the purposes and limitations of available assessment information. Different assessment strategies support different decision-making and/or accountability needs (see Webster, 1974). Purposes of testing include accounting to the public; appraising student achievement relative to specific instructional objectives; appraising student achievement relative to other pupils; assigning course grades; certifying the attainment of specific skills and knowledge; developing curriculum; diagnosing specific student strengths and weaknesses; evaluating experimental programs; grouping pupils for within-class instruction; helping students set goals; measuring student progress; placing students in special programs; planning educational programs; predicting student success in subsequent schooling or work; providing feedback to parents and students; and,

providing information for research into better approaches for helping students learn. Principals must be cognizant of why testing is being done and should plan their school testing program around specific information needs.

Basic Evaluation Concepts

Merely being familiar with basic measurement concepts and the use of test data for improving instruction is not sufficient for the modern educational leader. Pressure for greater accountability has caused many states and districts to develop accountability systems that, through a variety of methods, compare schools on variables such as student achievement, dropout, attendance, etc. (NCREL, 1993). Principals must be able to understand the basic concepts that provide for fair evaluation and comparison. When test data are being used for program evaluation they should be aware of the *Program Evaluation Standards* (1994) and when test data are being used in personnel evaluation they should be aware of the *Personnel Evaluation Standards* (1988).

This is a particularly difficult area since, according to a recent survey of graduate training programs in statistics, methodology, and measurement in psychology, even students who are being trained in psychology are not receiving training in the advanced measurement and statistical techniques that are required to assure fair comparisons among schools (Aiken, et. al., 1990). Nonetheless, principals should be conceptually familiar with certain rules of fair play.

Value-Added Methodology. While certainly not being required to understand the statistical techniques used in providing fair and appropriate value-added comparisons, principals should be aware that they should be held accountable for improvement, not for absolute unadjusted student test scores. They must understand that absolute test scores are as much a function of the students served as they are of the school's instructional program.

Influence of Background Factors on Learning. Student background factors such as ethnicity, gender, limited English proficiency status, socioeconomic status, and their interactions, impact learning. School level fairness variables include such things as student mobility, overcrowding conditions, average family income, average family education level, poverty index, percent students on free or reduced lunch, percent limited English proficient students, percent Black, Hispanic, and minority students, and percent teacher instructional days lost due to medical disability leave and unfilled vacancies. Principals should be aware of this and be aware

that there are statistical methods that adjust for these background variables, or, "level the playing field."

Accountability for Continuously Enrolled Students. Principals ought to be aware that their primary accountability should be for students that have been exposed to their instructional program, or continuously enrolled students.

The Importance of Multiple Outcome Variables. Principals should be cognizant of the fact that any accountability system should include multiple outcome variables.

Characteristics Of A Good Assessment Program

Finally, principals must be cognizant of the characteristics of a good assessment program so that they are able to raise questions and concerns when those characteristics are not present. Following are some criteria for judging a testing program.

Do the assessments agree with specified curricular objectives?

Before adopting a standardized assessment instrument or developing criterion-referenced assessment instruments, specifications should be examined in light of curricular objectives. In the absence of curricular objectives, assessment results are of limited use.

Are the assessments technically sound?

Measurement is only as good as the quality of the instruments used. Have the tests used in the assessment program been empirically validated? If they are standardized tests, data must be presented relative to characteristics of norm groups, item characteristics, reliability, and validity. If they are criterion-referenced tests, data must be presented relative to item characteristics and validity, as well as the instructional

objectives that the tests were designed to measure. If they are authenic assessments, data must be presented relative to validity and reliability as well as to the instructional objectives that the assessments are designed to measure.

Does the assessment program include all pupils?

If the assessment program is voluntary, or if it is given only to special groups of students, or if it is not given to certain groups of students, or if certain groups of students score at chance level, erroneous conclusions may be drawn about district achievement levels. More important, certain students may be educationally impaired because of lack of information for diagnostic and guidance purposes. Some characteristics of a comprehensive assessment program include:

- provision for functional level assessment,
- provision for assessment in primary language,
- provision for makeup assessment for absentees.

Are the assessments administered at regular intervals and are they well timed?

Are there regular assessment periods or are the tests administered in a haphazard manner? Without regular assessment periods, it is difficult to assess longitudinal pupil growth. If there are regular assessment periods, it is essential that they be timed so that test results can be maximally useful to the instruction process.

Are the assessments appropriately scaled?

In order for longitudinal comparisons to be valid, the assessments must be appropriately scaled. That is, scaling must be done across test levels and forms to assure comparability of results.

Are the assessments administered properly?

Poor test administration ruins the validity of assessment results. Oversights on the part of building test administrators, like failing to follow the standardized administration procedures outlined in most assessment manuals, failing to prevent extraneous interferences like fire drills and public address announcements during periods of assessment, grouping children in large groups so that they cannot hear instructions, failing to control the environment, etc., all detract from the validity of assessment results.

Are the assessments scored properly?

Test scoring is often a very complicated procedure. Scoring of performance assessments can be even more complicated. Quality control checks must be built in at every point in the scoring cycle. This is essential whether hand or machine scoring is employed. Care must also be taken to assure that the correct scores are related to the appropriate students in the appropriate schools and that, in the case of performance assessments, the correct scoring protocols are employed appropriately.

Are assessment results reported rapidly to teachers and counselors?

Assessment results obtained five months after the assessments were administered are of limited utility. Every effort should be

made to place interpretable results in the hands of teachers and counselors within two weeks of the assessment time.

Are assessment results used?

Assessment results must be reported in such a manner that they can be used by teachers and counselors. Both norm-referenced and criterion-referenced instruments and interpretation must be available. Constant checks must be made with teachers and counselors to determine unmet needs and expectations relative to assessment.

Is there a Staff Development Program on the use of assessment results?

Teachers and counselors must be educated so that they can make full use of assessment results. Detailed staff development programs must be designed so that every school has on its staff at least one person whose function is a thorough understanding of the assessment program. This person must then assume the responsibility of training the remainder of the faculty in the administration, use, and interpretation of assessment instruments.

Are assessment results reported accurately to parents and to the community?

Are appropriate comparisons made? Assessment results should not be released in isolation. They are only one indicator of the success of the schools. Are data reported that focus on improvement. Is appropriate value-added methodology used in reporting results?

Are there provisions for alternative assessments?

> Testing programs should include multiple indicators of student progress. Performance tests should be an integral part of instructional programs, even if done by teachers on a sampling basis. Portfolio analyses should be part of the accountability system.

Summary

This paper has attempted to outline some basic concepts in the areas of measurement and evaluation that principals must have in order to successfully operate a site-based managed school. While it is probably not realistic to expect educational administrators to be familiar with the *Standards for Educational and Psychological Testing* (1985), the *Program Evaluation Standards* (1994), and the *Personnel Evaluation Standards* (1988), these documents would provide excellent resources for the graduate training of educational administrators and would, if used, create an informed clientele for the use of educational tests in a number of different settings for a variety of purposes. Failing this, administrators must be at least conceptually familiar with several basic measurement concepts (reliability, validity, criterion-referenced testing, norm-referenced testing, scaling, performance testing, portfolio analysis, and the importance of test administration procedures); must be able to link curriculum content and instructional strategies to assessment information; must be conceptually familiar with several basic evaluation concepts (value-added, influence of background factors on learning, the importance of multiple outcome measures); and, must be cognizant of the characteristics of a good assessment program so that they may function as informed participants in the assessment and evaluation process.

References

Aiken, L. S., West, S. G., Sechrest, L., Reno, R. R. (1990) *Graduate Training In Statistics, Methodology, and Measurement In Psychology: A Survey of Ph.D. Programs In North America.* American Psychologist, 45, 6, 721-734

American Association of School Administrators (1993) *Professional Standards for the Superintendency.* AASA: Arlington, Virginia, 21-00423

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985) *Standards for Educational and Psychological Testing.* Washington, D. C.: American Psychological Association

Impara, J. C., Plake, B. S., and Fager, J. J. (1993a) *Educational Administrators' and Teachers' Knowledge of Classroom Assessment.* Journal of School Leadership, 3, 5, 510-521.

Impara, J. C., Plake, B. S., and Fager, J. J. (1993b) *Teachers' Assessment Background and Attitudes Toward Testing.* Theory Into Practice, 32, 2, 113-117.

Impara, J. C., Plake, B. S., and Merwin, J. (1994) *Student Assessment Tasks and Knowledge: Comparing Superintendents and Elementary and Secondary Principals.* Journal of School Leadership, 4, 5, 517-529

Mendro, R. L., Webster, W. J., Anderson, M., Babu, S., and Bearden, D. K. (1994) *Longitudinal Achievement Trends, 1981-94.* Dallas, Texas: Dallas Public Schools, REIS94-323

National Association of Elementary School Principals (1991) *Proficiencies for Principals, Elementary and Middle School, Revised.* NAESP: Alexandria, Virginia, ED 272 972

North Central Regional Educational Laboratory (1993) *Association of State Assessment Programs Annual Survey.* Council of Chief School Officers.

Popham, W. J., and Hambleton, R. K. (1990) *Can You Pass the Test on Testing?* Principal, 69, 3, 38-39.

The Joint Committee on Standards for Educational Evaluation (1988) *The Personnel Evaluation Standards*. Newbury Park: Sage Publications

The Joint Committee on Standards for Educational Evaluation (1994) *The Program Evaluation Standards*. Thousand Oaks: Sage Publications

Webster, W. J. (1974) *What's The Score On Testing?* Invited Address To The National School Board Association, Houston, Texas.